Computational Statistics
https://doi.org/10.1007/s00180-021-01183-3

ORIGINAL PAPER

RMSE-minimizing confidence intervals for the binomial parameter

Kexin Feng[1] · **Lawrence M. Leemis**[1] · **Heather Sasinowska**[1]

Received: 22 September 2020 / Accepted: 1 December 2021

Abstract

Let X be the number of successes in n mutually independent and identically distributed Bernoulli trials, each with probability of success p. For fixed n and α, there are $n + 1$ distinct two-sided $100(1-\alpha)\%$ confidence intervals for p associated with the outcomes $X = 0, 1, 2, \ldots, n$. There is no known exact non-randomized confidence interval for p. Existing approximate confidence interval procedures use a formula, which often requires numerical methods to implement, to calculate confidence interval bounds. The bounds associated with these confidence intervals correspond to discontinuities in the actual coverage function. The paper does not aim to provide a formula for the confidence interval bounds, but rather to select the confidence interval bounds that minimize the root mean square error of the actual coverage function for sample size n and significance level α in the frequentist context.

Keywords Actual coverage function · Approximate confidence interval · Binary data · Binomial distribution · Dyck word

1 Introduction

The binomial distribution has widespread applications in statistics. Its applications appear in public through news and surveys almost daily; therefore, it is not surprising that calculating an interval estimator for the binomial proportion has become a popular topic in statistics. Dozens of confidence interval procedures have been developed over the last 100 years that include applications such as Monte Carlo simulation, survey sampling, and survival analysis. In this paper we develop an algorithm for constructing an approximate two-sided $100(1-\alpha)\%$ confidence interval for the binomial proportion with actual coverage that is as close as possible to the nominal coverage (a.k.a. the stated coverage or the confidence level).

✉ Lawrence M. Leemis
leemis@math.wm.edu

[1] Department of Mathematics, William & Mary, Williamsburg, VA, USA

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli(p) population with unknown parameter p satisfying $0 < p < 1$. Let $X = \sum_{i=1}^{n} X_i$ be the number of successes. The maximum likelihood estimator for p is $\hat{p} = X/n$, which is an intuitive, unbiased, and consistent estimator of p. Including $p = 0$ and $p = 1$ in the parameter space includes two deterministic distributions for X with support values $X = 0$ and $X = n$, respectively. This situation seldom occurs in practice.

All of the confidence intervals developed to date are approximate, rather than exact confidence intervals. Each relies on a heuristic that results in a confidence interval for p having an actual coverage that is close to the nominal coverage. They provide a formula that gives the lower and upper bounds for an approximate confidence interval for p. Our approach here differs from all previous confidence interval procedures in that we choose confidence interval bounds that optimize a measure of performance known as the root mean square error (RMSE) of the actual coverage function in a frequentist context in which all values of p are equally weighted. As measured by the RMSE, our confidence intervals come closer to the nominal coverage $1 - \alpha$ than existing confidence intervals, and the percentage-wise improvement is particularly pronounced for smaller sample sizes.

The sequence of steps outlined in this paragraph include the contributions of this research. First, we define the *actual coverage function* as a piecewise function in which each piece is a portion of an *acceptance curve*. Second, we define the RMSE and establish minimizing the RMSE as the criterion for constructing a confidence interval for p. Third, we establish a one-to-one equivalence between the jumps from one acceptance curve to the next in the actual coverage function and a *symmetric Dyck path*. Fourth, we devise an algorithm for constructing RMSE-minimizing confidence intervals for p. Fifth, we discuss problems and their heuristic solutions concerning these confidence intervals. Sixth, we implement the algorithm in R and place the implementation on the comprehensive R archive network (CRAN).

Section 2 contains a brief literature review that defines (a) various classes of confidence intervals, (b) the actual coverage function, (c) acceptance curves, (d) the RMSE, and (e) Dyck paths. Section 3 concerns sample sizes $n = 1$ and $n = 2$ in which we manually calculate the RMSE-minimizing confidence interval bounds. Section 4 surveys several existing confidence interval procedures and compares them with the RMSE-minimizing confidence interval bounds for $n = 1$ and $n = 2$. Section 5 discusses the procedure for constructing the RMSE-minimizing confidence interval. Section 6 introduces "smoothness" as a preferable property of the binomial confidence intervals and a set of constraints imposed on the RMSE-minimizing confidence interval to achieve it. Section 7 illustrates the use of the RMSE-minimizing confidence interval in an application, and Sect. 8 contains conclusions.

2 Literature review

A comprehensive survey on statistical intervals in general and confidence intervals in particular is given by Meeker et al. (2017). We begin with a general classification of confidence intervals in the context of a random sample of n observations from a Bernoulli(p) population. The random variables L and U are known as the *lower bound* and the *upper bound* of the confidence interval. The probability $1-\alpha$ will be referred to here as the *nominal coverage*. The bounds L and U are functions of the sample size n, the nominal coverage $1-\alpha$, and the number of successes x. A random interval for the unknown parameter p of the form $L < p < U$ is

- an *exact* two-sided $100(1-\alpha)\%$ confidence interval for p provided
$$P(L < p < U) = 1-\alpha$$
for all values of p,
- an *approximate* two-sided $100(1-\alpha)\%$ confidence interval for p provided
$$P(L < p < U) \neq 1-\alpha$$
for some value of p,
- an *asymptotically exact* two-sided $100(1-\alpha)\%$ confidence interval for p provided
$$\lim_{n\to\infty} P(L < p < U) = 1-\alpha$$
for all values of p, and
- a *conservative* two-sided $100(1-\alpha)\%$ confidence interval for p provided
$$P(L < p < U) \geq 1-\alpha$$
for all values of p.

The probability $P(L < p < U)$ which appears in each of the four classes of confidence intervals defined above is known as the actual coverage of the confidence interval for a particular value of p. The *actual coverage function* $c(p)$ of a confidence interval for the binomial proportion is

$$c(p) = \sum_{x=0}^{n} I(x, p)\binom{n}{x} p^x(1-p)^{n-x},$$

where $I(x, p)$ is an indicator function that denotes whether a confidence interval includes the binomial proportion p when the number of successes $X = x$. As p is increased, the binomial probability terms in $c(p)$ will be added or removed from the summation at the confidence interval bounds.

The defining formula for the actual coverage function $c(p)$, and the fact that the lower bounds and upper bounds on *any* confidence interval procedure for the binomial proportion p are nondecreasing functions of x, means that the actual coverage function $c(p)$ must lie on one of the *acceptance curves* defined as

$$b(p, x_0, x_1) = \sum_{x=x_0}^{x_1} \binom{n}{x} p^x(1-p)^{n-x}$$

for a prescribed value of p satisfying $0 < p < 1$, and for integers x_0 and x_1 satisfying $0 \le x_0 \le x_1 \le n$. These acceptance curves are the sum of an uninterrupted sequence of probability mass function values from the binomial distribution. The values of p associated with the discontinuities in the actual coverage function are the confidence interval bounds. The discontinuities in $c(p)$ are a result of either an increase in x_0 or an increase in x_1 in $b(p, x_0, x_1)$. If x_0 is increased, the discontinuity is associated with an upper confidence interval bound; if x_1 is increased, the discontinuity is associated with a lower confidence interval bound. In general, there are $2n + 1$ segments in the actual coverage function, which are associated with $2n+2$ confidence interval bounds. There is no exact confidence interval for the binomial proportion p from a random sample of n binary data values because the actual coverage function for all confidence interval procedures must transition between these acceptance curves.

We now define measures of performance associated with a confidence interval for the binomial proportion p. First, the mean actual coverage m for a confidence interval procedure is the average value of the actual coverage function for fixed n and α:

$$m = \int_0^1 c(p)\,dp.$$

The variance of the actual coverage v is defined as

$$v = \int_0^1 c^2(p)\,dp - m^2.$$

The two measures of performance can be combined into a single measure by devising a calculation that is similar to the root mean squared error (that is, the square root of the variance plus the squared bias):

$$\text{RMSE} = \sqrt{v + (m - (1 - \alpha))^2},$$

as defined by Park and Leemis (2019). Our goal in this paper is to devise a confidence interval for p that minimizes the RMSE. Minimizing the RMSE makes the actual coverage function of the approximate confidence interval for p come as close as possible to the nominal coverage. Other measures of performance could be used; in fact, the mean absolute deviation would result in identical confidence intervals. Minimizing the RMSE implies that the confidence interval is entirely in the frequentist context; all values of p are equally weighted.

There is a way to calculate m and v that avoids numerical integration. For a fixed sample size n, a confidence interval procedure for the binomial proportion p associated with $x = 0, 1, 2, \ldots, n$ successes results in $n + 1$ confidence intervals. Thus, there are $2n + 2$ associated confidence interval bounds. Let $p_1, p_2, \ldots, p_{2n+2}$ denote these ordered confidence interval bounds. These bounds correspond to the endpoints of the piecewise actual coverage function $c(p)$. Each of the $2n+1$ pieces of $c(p)$ corresponds to a piece of one of the acceptance curves $b(p, x_0, x_1)$. Let x_{0i} and x_{1i} denote the lower and upper summation limits associated with the ith piece of $c(p)$, for $i =$

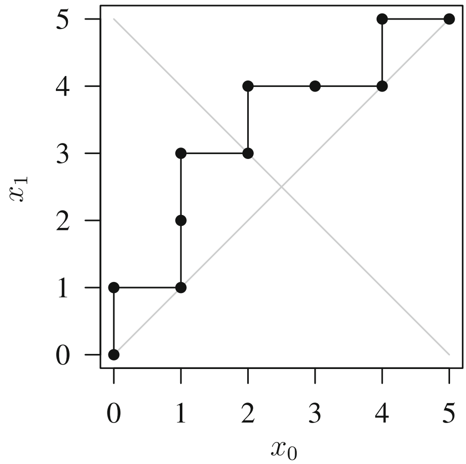

Fig. 1 A symmetric Dyck path for $n = 5$

$1, 2, \ldots, 2n+1$. Using this notation and the binomial theorem, an expression for the mean actual coverage which avoids numerical integration is (Park and Leemis 2019)

$$m = \int_0^1 c(p)dp = \sum_{i=1}^{2n+1} \sum_{x=x_{0i}}^{x_{1i}} \binom{n}{x} \sum_{k=0}^{n-x} \binom{n-x}{k} (-1)^k \left[\frac{p_{i+1}^{k+x+1} - p_i^{k+x+1}}{k+x+1} \right].$$

This derivation exploits the fact that the actual coverage function is a piecewise polynomial function in p which has a closed-form integration. Using a similar approach and again applying the binomial theorem, an expression for the variance of the actual coverage which avoids numerical integration is

$$v = \left\{ \sum_{i=1}^{2n+1} \sum_{x=x_{0i}}^{x_{1i}} \sum_{y=x_{0i}}^{x_{1i}} \binom{n}{x}\binom{n}{y} \sum_{k=0}^{2n-x-y} \binom{2n-x-y}{k} (-1)^k \left[\frac{p_{i+1}^{k+x+y+1} - p_i^{k+x+y+1}}{k+x+y+1} \right] \right\} - m^2.$$

We now introduce a symmetric Dyck word and the associated symmetric Dyck path. Kása (2009) provides a definition for a symmetric Dyck word. Let $B = \{0, 1\}$ be a binary alphabet, a discrete set of two symbols, and a word $x_1 x_2 \ldots x_n \in B^n$. Let $h : B \longrightarrow \{-1, 1\}$ be a valuation function with $h(0) = 1$, $h(1) = -1$, and $h(x_1 x_2 \ldots x_n) = \sum_{i=1}^{n} h(x_i)$. A word $X = x_1 x_2 \ldots x_{2n} \in B^{2n}$ is called a *Dyck word* if it satisfies the following conditions:

$$h(x_1 x_2 \ldots x_i) \geq 0, \text{ for } i = 1, 2, \ldots, 2n-1 \quad \text{and} \quad h(x_1 x_2 \ldots x_{2n}) = 0.$$

A *symmetric Dyck word* satisfies $(1 - x_1)(1 - x_2) \ldots (1 - x_{2n-1})(1 - x_{2n}) = x_{2n} x_{2n-1} \ldots x_2 x_1$.

A *symmetric Dyck path* is a staircase walk from $(0, 0)$ to (n, n) in the (x_0, x_1) plane, which lies strictly above the line $x_1 = x_0$ and is symmetric with respect to the line $x_1 = n - x_0$. If we associated 0 with an upward step and 1 with a rightward step, we could easily convert a symmetric Dyck word into a symmetric Dyck path. Figure 1 illustrates a symmetric Dyck path of length 10. It corresponds to the symmetric Dyck word 0100101101. The first entry is 0 in the Dyck word, so the first step is from (0, 0) to (0, 1). The second entry is 1 in the Dyck word, so the second step is from (0, 1) to (1, 1).

There is a one-to-one relationship between the jumps between the acceptance curves in the actual coverage function for a sample size n and symmetric Dyck paths of length $2n$. Each discontinuity in the actual coverage function is a result of an increase in either x_0 or x_1. When x_0 is increased (a *rightward* move in the symmetric Dyck path), a binomial probability mass function value is *dropped* from the summation in the definition of $b(p, x_0, x_1)$, so the actual coverage function takes a *downward* step. When x_1 is increased (an *upward* move in the symmetric Dyck path), a binomial probability mass function value is *added* to the summation in the definition of $b(p, x_0, x_1)$, so the actual coverage function takes an *upward* step. Given the sequence of (x_0, x_1) pairs associated with an actual coverage function, we can construct a symmetric Dyck path. Conversely, given a symmetric Dyck path of length $2n$, we are able to generate the sequence of acceptance curves associated with an actual coverage function of order n. The number of symmetric Dyck paths of order n is

$$\binom{n}{\lceil n/2 \rceil},$$

which is the n^{th} central binomial coefficient (Deng et al. 2009).

3 Small sample calculations

To better understand the behavior of the actual coverage function and the intuition and geometry behind the RMSE-minimizing confidence interval, we manually calculate the optimal confidence interval bounds for $n = 1$ and $n = 2$ in the next two subsections. Even though the parameter space is $0 < p < 1$, the lower confidence interval bound is 0 when $x = 0$ and the upper confidence interval bound is 1 when $x = n$.

3.1 One-sample case

There is only one possible symmetric Dyck path for $n = 1$. The path starts at (0, 0), moves to (0, 1), and ends at (1, 1). It is illustrated in Fig. 2.

The three acceptance curves corresponding to the symmetric Dyck path are

$$\begin{aligned} b(p, 0, 0) &= 1 - p, \\ b(p, 0, 1) &= 1, \\ b(p, 1, 1) &= p, \end{aligned}$$

for $0 < p < 1$. Since the symmetric Dyck path goes from $(x_0, x_1) = (0, 0)$ to (0, 1) to (1, 1), the actual coverage function will jump from $b(p, 0, 0)$ to $b(p, 0, 1)$ to $b(p, 1, 1)$. This is an illustration of the one-to-one correspondence between a symmetric Dyck path and a sequence of jumps between acceptance curves. Since there are two confidence intervals for $n = 1$, one associated with $x = 0$ and the other associated with

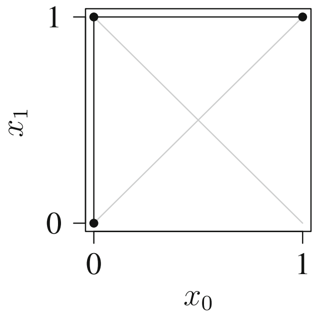

Fig. 2 Symmetric Dyck path for $n = 1$

$x = 1$, there will be two discontinuities in the actual coverage function on $0 < p < 1$. We use p_1 and p_2 to denote the ordered unknown confidence interval bounds. Since the symmetric Dyck path moves from (0, 0) to (0, 1) initially, p_1 is associated with a lower bound. Since the symmetric Dyck path then moves from (0, 1) to (1, 1), p_2 is associated with an upper bound. So the two confidence intervals are $0 < p < p_2$ for $x = 0$ and $p_1 < p < 1$ for $x = 1$. First, we calculate the mean actual coverage m:

$$\begin{aligned} m &= \int_0^1 c(p)\,dp \\ &= \int_0^{p_1} (1-p)\,dp + \int_{p_1}^{p_2} 1\,dp + \int_{p_2}^1 p\,dp \\ &= -\frac{1}{2}p_1^2 + p_2 + \frac{1}{2} - \frac{1}{2}p_2^2. \end{aligned}$$

Next, we calculate the variance of the actual coverage v:

$$\begin{aligned} v &= \int_0^1 c^2(p)\,dp - m^2 \\ &= \int_0^{p_1} (1-p)^2\,dp + \int_{p_1}^{p_2} 1^2\,dp + \int_{p_2}^1 p^2\,dp - m^2 \\ &= p_1 - p_1^2 + \frac{1}{3}p_1^3 + p_2 - p_1 + \frac{1}{3} - \frac{1}{3}p_2^3 - \left(-\frac{1}{2}p_1^2 + p_2 + \frac{1}{2} - \frac{1}{2}p_2^2\right)^2 \\ &= \frac{1}{3}p_1^3 - \frac{1}{2}p_1^2 + \frac{1}{12} + \frac{2}{3}p_2^3 - \frac{1}{4}p_1^4 - \frac{1}{4}p_2^4 - \frac{1}{2}p_1^2p_2^2 + p_1^2p_2 - \frac{1}{2}p_2^2. \end{aligned}$$

We arbitrarily choose $\alpha = 0.05$, which results in the mean square error

$$\begin{aligned} \text{RMSE}^2 = v + \left(m - \frac{19}{20}\right)^2 &= \frac{1}{3}p_1^3 - \frac{1}{20}p_1^2 + \frac{1}{12} \\ &+ \left(\frac{19}{20}\right)^2 - \frac{9}{10}p_2 + \frac{1}{4} - \frac{19}{20} - \frac{1}{3}p_2^3 + \frac{19}{20}p_2^2. \end{aligned}$$

Table 1 RMSE-minimizing confidence interval bounds for $n = 1$

	L	U
$x = 0$	0	0.9
$x = 1$	0.1	1

In order to minimize the mean square error (which also minimizes the RMSE), we take partial derivatives with respect to p_1 and p_2 and set them equal to 0:

$$\frac{\partial \text{RMSE}^2}{\partial p_1} = p_1^2 - \frac{1}{10} p_1 = 0$$
$$\frac{\partial \text{RMSE}^2}{\partial p_2} = -\frac{9}{10} - p_2^2 + \frac{19}{10} p_2 = 0.$$

Solving this simultaneous set of equations for p_1 and p_2 results in

$$p_1 = \frac{1}{10}, \quad p_2 = \frac{9}{10}.$$

We did not exploit the fact that $p_2 = 1 - p_1$ in this derivation. Replacing the arbitrary designation of "success" with "failure" in each Bernoulli trial forces this relationship between lower and upper bounds. For general n, $p_{2n+1-i} = 1 - p_i$ always holds due to the symmetry of the binomial confidence interval for $i = 1, 2, \ldots, 2n$.

The confidence intervals for $n = 1$ are displayed in Table 1. The actual coverage function is shown in Fig. 3 by the solid lines. The three acceptance curves are shown in gray. The actual coverage function has a discontinuity wherever the vertical distance between the current acceptance curve and the nominal coverage $1 - \alpha$ becomes greater than the vertical distance between the subsequent acceptance curve (as defined by the symmetric Dyck path) and the nominal coverage $1 - \alpha$.

The values of m, v, and RMSE for these approximate 95% confidence intervals are straightforward to calculate in this case. The value of m is

$$\begin{aligned} m &= \int_0^1 c(p)\, dp \\ &= \int_0^{0.1} (1-p)\, dp + \int_{0.1}^{0.9} 1\, dp + \int_{0.9}^{1} p\, dp \\ &= 0.99. \end{aligned}$$

The value of v is

$$\begin{aligned} v &= \int_0^1 c^2(p)\, dp - m^2 \\ &= \int_0^{0.1} (1-p)^2\, dp + \int_{0.1}^{0.9} 1\, dp + \int_{0.9}^{1} p^2\, dp - m^2 \\ &= \frac{17}{30{,}000}. \end{aligned}$$

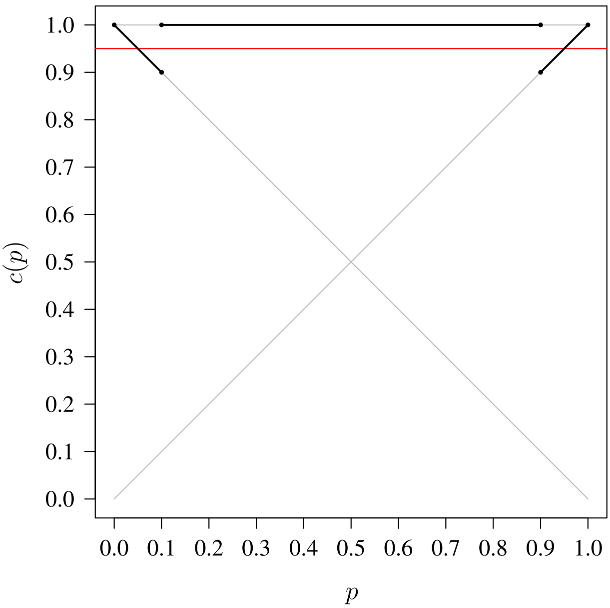

Fig. 3 Actual coverage function for $n = 1$ and $\alpha = 0.05$

Finally, the value of the RMSE is

$$\begin{aligned} \text{RMSE} &= \sqrt{v + \big(m - (1-\alpha)\big)^2} \\ &= \sqrt{\frac{17}{30{,}000} + \left(\frac{99}{100} - \frac{95}{100}\right)^2} \\ &= \sqrt{\frac{65}{30{,}000}} \\ &\cong 0.047. \end{aligned}$$

3.2 Two-sample case

The situation for sample size $n = 2$ is more complicated. There are a total of $\binom{2}{1} = 2$ symmetric Dyck paths from $(x_0, x_1) = (0, 0)$ to $(x_0, x_1) = (2, 2)$, which are shown in Fig. 4. We consider both paths separately and compare their RMSEs.

3.2.1 Symmetric Dyck path 1

Of the six potential acceptance curves associated with $n = 2$, the five acceptance curves corresponding to the first symmetric Dyck path are

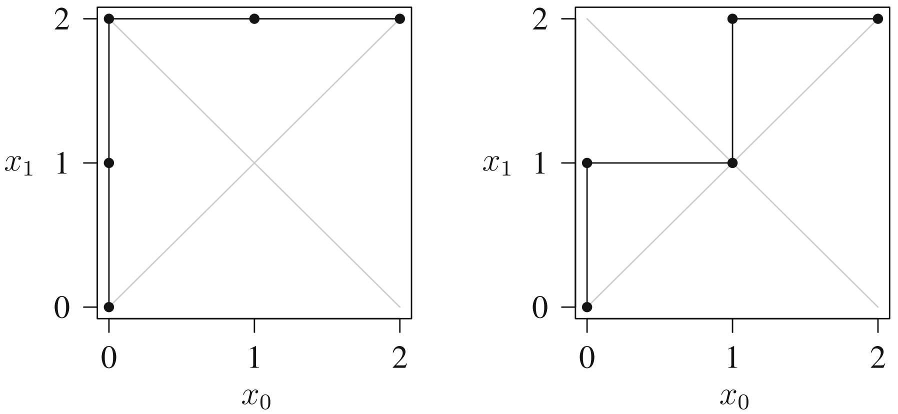

Fig. 4 Symmetric Dyck paths for $n = 2$: Path 1 (left) and Path 2 (right)

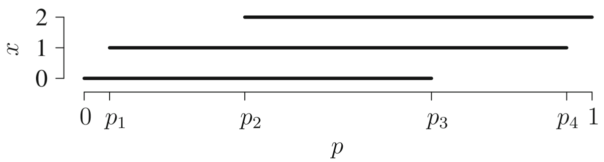

Fig. 5 RMSE-minimizing confidence intervals for $n = 2$ and $\alpha = 0.05$

$$\begin{aligned}
b(p, 0, 0) &= (1 - p)^2, \\
b(p, 0, 1) &= (1 - p)^2 + 2p(1 - p) = 1 - p^2, \\
b(p, 0, 2) &= 1, \\
b(p, 1, 2) &= 2p - p^2, \\
b(p, 2, 2) &= p^2,
\end{aligned}$$

for $0 < p < 1$. There are a total of $2 \times 2 + 1 = 5$ segments in the actual coverage functions, which means there are $2 \times 2 + 2 = 6$ confidence interval bounds, including $p = 0$ and $p = 1$. There are four unknown confidence interval bounds, denoted by p_1, p_2, p_3, and p_4. The three confidence intervals associated with the confidence interval bounds are

$$\begin{aligned}
x = 0 &\quad\Rightarrow\quad 0 < p < p_3 \\
x = 1 &\quad\Rightarrow\quad p_1 < p < p_4 \\
x = 2 &\quad\Rightarrow\quad p_2 < p < 1.
\end{aligned}$$

Notice that $p_4 = 1 - p_1$ and $p_3 = 1 - p_2$ by symmetry. We know that $0 < p_1 < p_2 < p_3 < p_4 < 1$, as illustrated by the confidence intervals depicted in Fig. 5.

Table 2 RMSE-minimizing confidence interval bounds for $n = 2$, Path 1

	L	U
$x = 0$	0	0.684
$x = 1$	0.050	0.950
$x = 2$	0.316	1

As in the previous subsection, we first calculate m:

$$\begin{aligned} m &= \int_0^1 c(p)\,dp \\ &= \int_0^{p_1} (1-p)^2\,dp + \int_{p_1}^{p_2} \left(1-p^2\right)\,dp + \int_{p_2}^{p_3} 1\,dp \\ &\quad + \int_{p_3}^{p_4} \left(2p-p^2\right)\,dp + \int_{p_4}^{1} p^2\,dp \\ &= \frac{2}{3}p_1^3 - p_1^2 - \frac{1}{3}p_2^3 + p_3 + p_4^2 - \frac{2}{3}p_4^3 + \frac{1}{3}p_3^3 - p_3^2 + \frac{1}{3}. \end{aligned}$$

Proceeding in the same fashion as in the case of $n = 1$, we calculate v and RMSE as before. In order to minimize the mean square error (which also minimizes the RMSE), we take partial derivatives with respect to p_1, p_2, p_3, and p_4.

$$\begin{aligned} \frac{\partial \text{RMSE}^2}{\partial p_1} &= -4p_1^3 + \frac{21}{5}p_1^2 - \frac{1}{5}p_1 = 0 \\ \frac{\partial \text{RMSE}^2}{\partial p_2} &= p_2^4 - \frac{1}{10}p_2^2 = 0 \\ \frac{\partial \text{RMSE}^2}{\partial p_3} &= -p_3^4 + 4p_3^3 - \frac{59}{10}p_3^2 + \frac{19}{5}p_3 - \frac{9}{10} = 0 \\ \frac{\partial \text{RMSE}^2}{\partial p_4} &= -4p_4^3 + \frac{39}{5}p_4^2 - \frac{19}{5}p_4 = 0. \end{aligned}$$

Solving this 4×4 simultaneous set of equations for p_1, p_2, p_3, and p_4 results in

$$p_1 = \frac{1}{20}, \quad p_2 = \frac{\sqrt{10}}{10}, \quad p_3 = 1 - \frac{\sqrt{10}}{10}, \quad p_4 = \frac{19}{20}.$$

The confidence intervals for $n = 2$ for this particular symmetric Dyck path are displayed in Table 2 and illustrated in Fig. 5. The associated actual coverage function for this path is shown in Figure 6.

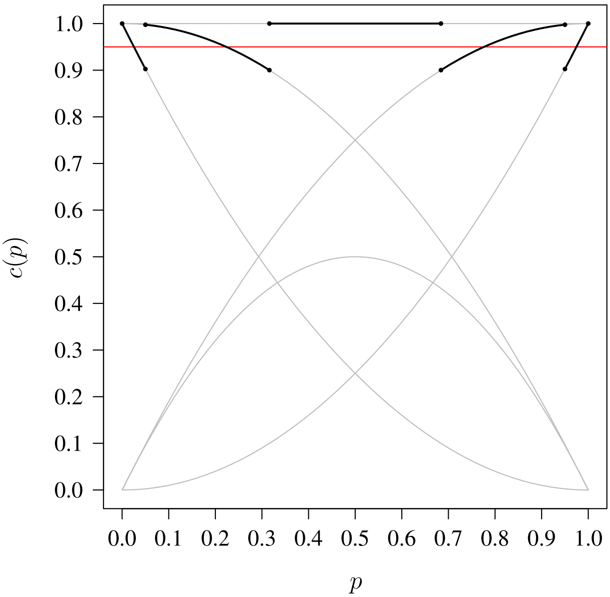

Fig. 6 Actual coverage function (black lines) for $n = 2$ and $\alpha = 0.05$, Path 1

3.2.2 Symmetric Dyck path 2

Of the six potential acceptance curves associated with $n = 2$, the five acceptance curves corresponding to the second symmetric Dyck path in Fig. 4 are

$$\begin{aligned} b(p, 0, 0) &= (1-p)^2, \\ b(p, 0, 1) &= (1-p)^2 + 2p(1-p) = 1 - p^2, \\ b(p, 1, 1) &= 2p(1-p), \\ b(p, 1, 2) &= 2p - p^2, \\ b(p, 2, 2) &= p^2, \end{aligned}$$

for $0 < p < 1$. We again denote the unknown confidence interval bounds by p_1, p_2, p_3, and p_4. The value of p_1 is the same as Path 1 because the first two segments are the same in the two paths. The calculation for the second confidence interval bound is unusual. If we take the partial derivative with respect to p_2, that is,

$$\frac{\partial \text{RMSE}^2}{\partial p_2} = -3p_2^4 + 8p_2^3 - \frac{79}{10}p_2^2 + \frac{19}{5}p_2 - \frac{9}{10},$$

the only solution is $p_2 = 1$, which does not satisfy $0 < p_2 < 1$ and results in the confidence interval $1 < p < 1$.

This unusual situation prompts us to consider a very brief "dwell time" (a term we will formally define in Sect. 5 associated with an ϵ-jump) on an acceptance curve.

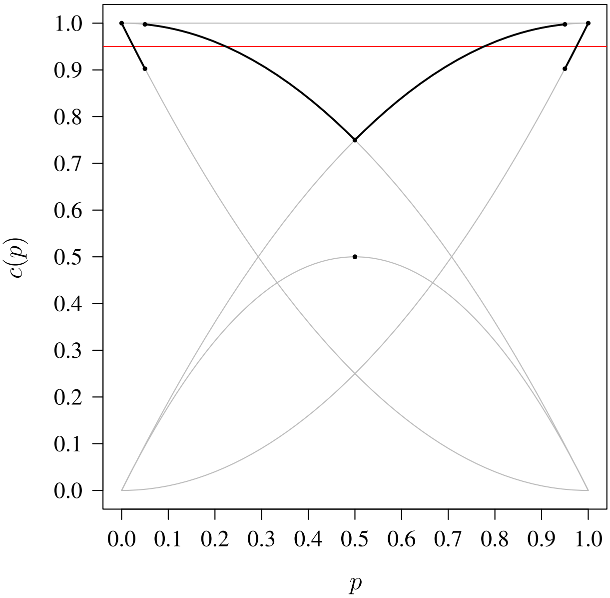

Fig. 7 Actual coverage function (black lines) for $n = 2$ and $\alpha = 0.05$, Path 2.

Since we are unable to find a point on $b(p, 1, 1)$ that minimizes the RMSE via calculus, the RMSE-minimizing procedure transitions from the acceptance curve $b(p, 0, 1)$ to $b(p, 1, 1)$, where it remains for only an instant, and then transitions to $b(p, 1, 2)$. Figure 7 illustrates this ϵ-jump at $p = 0.5$.

A visual inspection of Figs. 6 and 7 reveals that symmetric Dyck Path 1 results in a smaller RMSE than symmetric Dyck Path 2, so its confidence interval bounds comprise the RMSE-minimizing confidence interval.

4 Existing confidence interval procedures

To illustrate the geometry associated with an actual coverage function, acceptance curves, and symmetric Dyck paths for larger values of n, we consider the conservative Clopper–Pearson confidence interval (Clopper and Pearson 1934) for the binomial proportion p, which can be expressed as quantiles of the beta distribution:

$$B_{x,n-x+1,1-\alpha/2} < p < B_{x+1,n-x,\alpha/2},$$

for $x = 0, 1, 2, \ldots, n$, where the first two subscripts are the parameters of the beta distribution and the third subscript is a right-hand tail probability. Figure 8 contains three graphs that are associated with sample size $n = 10$ and nominal coverage $1 - \alpha = 1 - 0.05 = 0.95$ for the Clopper–Pearson confidence interval procedure. The top graph contains the acceptance curves in gray, the nominal coverage as a red horizontal line, and the actual coverage function as solid black lines. Whether the actual coverage function is right-continuous or left-continuous is not relevant in this paper, so both endpoints are given as solid circles. The middle graph shows the $n + 1 = 11$ possible confidence intervals associated with $x = 0, 1, 2, \ldots, 10$. The bottom graph shows the progression of x_0 and x_1 along the symmetric Dyck path associated with the jumps from one acceptance curve to another.

Notice that the confidence interval associated with $x = 1$ in Fig. 8, which is

$$0.003 < p < 0.445,$$

has an upper bound which is quite close to the lower bound of the confidence interval associated with $x = 8$, which is

$$0.444 < p < 0.975.$$

This is reflected in the graphs by an actual coverage function that has a very small dwell time on a particular acceptance curve of 0.001 on the top graph between $p = 0.444$ and $p = 0.445$ on the curve that is associated with the transition from $(x_0, x_1) = (1, 7)$ to $(x_0, x_1) = (1, 8)$ to $(x_0, x_1) = (2, 8)$. This does not pose any difficulty to the confidence intervals, as indicated in the middle graph. The fact that the upper bound associated with $x = 1$ is close to the lower bound for $x = 8$ is coincidental. However, cases will arise later in which these small dwell times do indeed cause difficulties with the confidence intervals.

Since the purpose of this paper is to construct confidence intervals with a minimal RMSE value, we will henceforth exclude conservative confidence intervals like the Clopper–Pearson or Blaker confidence intervals (Blaker 2000), although we mention in passing that the Blaker confidence interval has a smaller value of m than the associated Clopper–Pearson confidence interval for fixed values of n and α. Conservative confidence intervals have actual coverage which is always greater than or equal to the nominal coverage, so their RMSE values tend to be greater than those of non-conservative intervals. We end this literature review by briefly surveying four popular non-conservative confidence intervals which will play a role in the algorithm developed here.

The Wilson–score $100(1-\alpha)\%$ confidence interval for p has bounds (Wilson 1927)

$$\frac{1}{1 + z_{\alpha/2}^2/n} \left[\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}} \right],$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ percentile of the standard normal distribution. The Jeffreys $100(1-\alpha)\%$ confidence interval for p is a Bayesian credible interval that uses a Jeffreys noninformative prior distribution for p. The bounds of the Jeffreys confidence interval for p are percentiles of a beta random variable:

$$B_{x+1/2, n-x+1/2, 1-\alpha/2} < p < B_{x+1/2, n-x+1/2, \alpha/2}$$

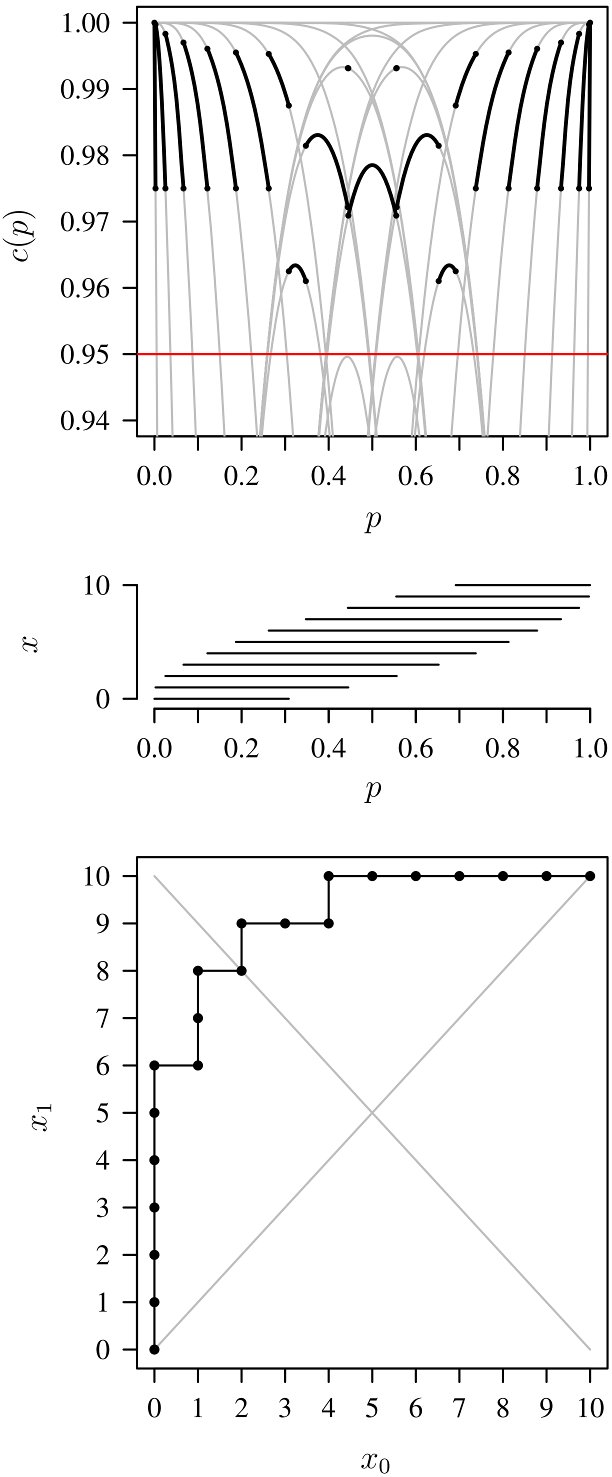

Fig. 8 Clopper–Pearson confidence intervals for $n = 10$ and $\alpha = 0.05$

for $x = 1, 2, \ldots, n-1$. When $x = 0$, the lower bound is set to zero and the upper is bound calculated using the formula above; when $x = n$, the upper bound is set to one and the lower bound is calculated using the formula above. The arcsine transformation uses a variance-stabilizing transformation when constructing a confidence interval for p. Using a modification suggested by Anscombe (1956), the bounds on a $100(1-\alpha)\%$ confidence interval for p are

$$\sin^2\left(\arcsin\left(\sqrt{\tilde{p}}\right) \pm \frac{z_{\alpha/2}}{2\sqrt{n}}\right),$$

where $\tilde{p} = (x + 3/8)/(n + 3/4)$. In the rare cases in which a confidence interval does not include the point estimator, one of the bounds is adjusted to include the point estimator. The bounds of the Agresti–Coull $100(1-\alpha)\%$ confidence interval, which was originally developed to approximate the Wilson-score confidence interval, are (Agresti and Coull 1998)

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}},$$

where $\tilde{n} = n + z^2_{\alpha/2}$ and $\tilde{p} = (x + z^2_{\alpha/2}/2)/\tilde{n}$.

These are not the only confidence intervals for p. Brown et al. (2001), for example, give a long list of binomial confidences intervals, including the logit interval, the likelihood ratio interval, and the Bayesian highest posterior density interval. Confidence intervals for p developed more recently include those by Balch (2020), Kim and Jang (2021), Lyles et al. (2020), Wilcox (2020), and Yaacoub et al. (2019). We consider only the four non-conservative confidence intervals described above in our confidence interval procedure because of their popularity and available implementation in software. The confidence intervals reviewed in this section tend to have poor performance for small n and values of p near 0 or 1. The goal of this paper is to find an approximate confidence interval for p whose actual coverage is as close as possible to the nominal coverage as measured by the RMSE.

Table 3 contains the confidence interval bounds for the four non-conservative confidence interval procedures and the RMSE-minimizing procedure for $n = 1$ and $\alpha = 0.05$. Three observations from the confidence intervals and RMSE values given in Table 3 are: (1) the RMSE-minimizing confidence interval procedure indeed produces the confidence interval with the lowest RMSE for $n = 1$, as it should; (2) the binomial confidence intervals are symmetric, as they should be; and (3) the RMSE is higher for the four existing confidence intervals because they are narrower (Wilson–score, Jeffreys, and Agresti–Coull) or wider (Arcsine) than the associated RMSE-minimizing confidence interval.

Table 4 gives the ordered confidence interval limits and the RMSE values associated with $n = 2$ and $\alpha = 0.05$. Again, the RMSE-minimizing confidence interval achieves the lowest RMSE. Some intuition can be gleaned from a plot of the actual coverage function for the Wilson–score confidence interval, which has the highest RMSE. Figure 9 shows that the actual coverage function for the Wilson–score confidence interval

Table 3 Confidence interval bounds and RMSEs for $n = 1$ and $\alpha = 0.05$

	Confidence interval bounds				RMSE
		p_1	p_2		
Wilson–score	0	0.207	0.793	1	0.0641
Jeffreys	0	0.147	0.853	1	0.0495
Arcsine	0	0.012	0.988	1	0.0499
Agresti–Coull	0	0.167	0.833	1	0.0532
RMSE-minimizing	0	0.1	0.9	1	0.0465

Table 4 Confidence interval bounds and RMSEs for $n = 2$ and $\alpha = 0.05$

	Confidence interval bounds						RMSE
		p_1	p_2	p_3	p_4		
Wilson–score	0	0.095	0.342	0.658	0.905	1	0.0461
Jeffreys	0	0.061	0.333	0.667	0.939	1	0.0392
Arcsine	0	0.009	0.230	0.770	0.991	1	0.0440
Agresti–Coull	0	0.095	0.290	0.710	0.905	1	0.0460
RMSE-minimizing	0	0.050	0.316	0.684	0.950	1	0.0387

procedure deviates more from the nominal coverage than the associated actual coverage function for the RMSE-minimizing confidence interval procedure depicted in Fig. 6.

5 RMSE-minimizing confidence interval

In principle, the procedure outlined in the previous section for determining RMSE-minimizing confidence interval bounds for $n = 1$ and $n = 2$ can be carried out for larger values of n. We modify an algorithm by Kása (2009) to generate all symmetric Dyck paths. The algorithm generates the first half of each symmetric Dyck path, and then the second half of the Dyck path is easily appended due to the symmetry. CPU time limitations will come into play for significantly larger values of n as the number of Dyck paths that need to be inspected grows on the order of the factorial of n. In this section, we introduce the steps associated with calculating the bounds of an RMSE-minimizing two-sided $100(1-\alpha)\%$ confidence interval for p and associated problems that arise in the design of the algorithm.

Consider the general case of an arbitrary sample size n. Let $p_1, p_2, \ldots, p_{2n}$ denote the ordered confidence interval bounds on $0 < p < 1$, and $b_1(p), b_2(p), \ldots, b_{2n+1}(p)$ denote the acceptance curves associated with one particular Dyck path. We have suppressed the last two arguments, x_0 and x_1, on b for compactness and to simplify the notation. The mean square error is

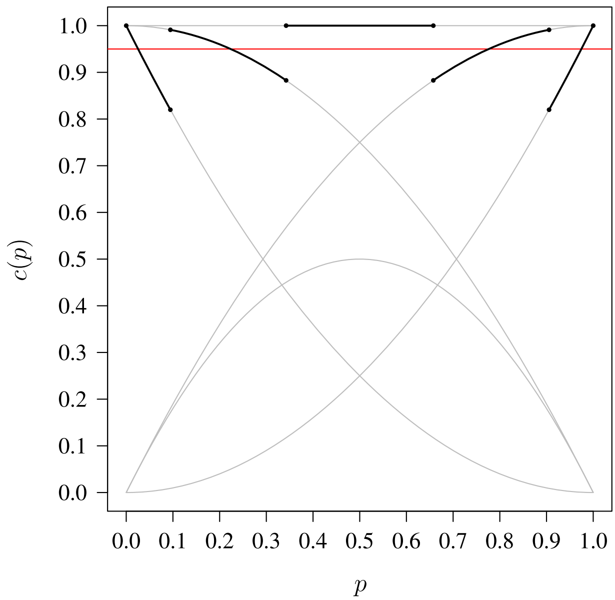

Fig. 9 Actual coverage function for Wilson–score interval for $n = 2$ and $\alpha = 0.05$

$$
\begin{aligned}
\mathrm{RMSE}^2 &= v + \big(m - (1-\alpha)\big)^2 \\
&= \int_0^1 c^2(p)\,dp - \left[\int_0^1 c(p)\,dp\right]^2 + \left(\int_0^1 c(p)\,dp - (1-\alpha)\right)^2 \\
&= \int_0^1 c^2(p)\,dp - 2(1-\alpha)\int_0^1 c(p)\,dp + (1-\alpha)^2 \\
&= \int_0^{p_1} b_1^2(p)\,dp + \int_{p_1}^{p_2} b_2^2(p)\,dp + \cdots + \int_{p_{2n}}^1 b_{2n+1}^2(p)\,dp \\
&\quad -2(1-\alpha)\left(\int_0^{p_1} b_1(p)\,dp + \int_{p_1}^{p_2} b_2(p)\,dp + \cdots + \int_{p_{2n}}^1 b_{2n+1}(p)\,dp\right) + (1-\alpha)^2.
\end{aligned}
$$

For all values of n, the first acceptance curve, $b_1(p) = (1-p)^n$, corresponds to $(x_0, x_1) = (0, 0)$, and the second acceptance curve, $b_2(p) = (1-p)^n + np(1-p)^{n-1}$, corresponds to $(x_0, x_1) = (0, 1)$. When we calculate p_1, we only need to consider the terms in RMSE^2 that contain p_1. In order to minimize the mean square error (which also minimizes the RMSE), we take the partial derivative with respect to p_1:

$$
\begin{aligned}
\frac{\partial \mathrm{RMSE}^2}{\partial p_1} &= (1-p_1)^{2n} - (1-p_1)^{2n} - 2np_1(1-p_1)^{2n-1} - n^2 {p_1}^2(1-p_1)^{2n-2} - \\
&\qquad 2(1-\alpha)(1-p_1)^n + 2(1-\alpha)(1-p_1)^n + 2(1-\alpha)np_1(1-p_1)^{n-1} \\
&= -2np_1(1-p_1)^{2n-1} - n^2 {p_1}^2(1-p_1)^{2n-2} + 2(1-\alpha)np_1(1-p_1)^{n-1} \\
&= np_1\left[2(1-\alpha)(1-p_1)^{n-1} - 2(1-p_1)^{2n-1} - np_1(1-p_1)^{2n-2}\right].
\end{aligned}
$$

Equating this partial derivative to zero, p_1 can be found by numerically solving

$$2(1-p_1)^n + np_1(1-p_1)^{n-1} - 2(1-\alpha) = 0.$$

We can apply a similar calculation to any confidence interval bound. To obtain the value of p_c that minimizes the RMSE, for $c = 1, 2, \ldots, n$, we only need to consider the terms in RMSE^2 that contain p_c. More specifically, we will focus on the terms containing $b_c(p)$ and $b_{c+1}(p)$. Then we will take the partial derivative respect to p_c, set it equal to 0, and solve for p_c. In this way, we derive a general formula, solvable for confidence interval bounds.

Let $0, p_1, p_2, \ldots, p_{2n}, 1$ be the sorted $2n+2$ endpoints of the confidence intervals. Proceeding in the same fashion as the derivation of the value of p_1 above, the value of p_c, for $c = 1, 2, \ldots, n$, which minimizes the RMSE can be determined by numerically solving

$$b_c(p_c) + b_{c+1}(p_c) = 2(1-\alpha).$$

However, this formula can only be applied when there exists a solution between 0 and 1. In many cases there is no such solution, which leads to the discussion of the dwell time and ϵ-jumps. By symmetry, $p_c = 1 - p_{2n-c+1}$, for $c = n+1, n+2, \ldots, 2n$.

For a particular confidence interval procedure for p with fixed parameters α and n, for positive integer n and $0 < \alpha < 1$, the *dwell time* on an acceptance curve associated with fixed summation limits (x_0, x_1) is the difference between the values of p between two adjacent discontinuities of the actual coverage function on that acceptance curve (including $p = 0$ and $p = 1$).

For the Clopper–Pearson 95% confidence interval for $n = 10$ illustrated in Fig. 8, for example, the dwell time on the acceptance curve associated with $(x_0, x_1) = (0, 0)$ is $0.0025 - 0.0000 = 0.0025$. The dwell time on the acceptance curve associated with $(x_0, x_1) = (0, 1)$ is $0.0252 - 0.0025 = 0.0227$. The longest dwell time for the Clopper–Pearson 95% confidence interval is associated with $(x_0, x_1) = (2, 8)$ is $0.5550 - 0.4450 = 0.1100$. The smallest dwell times for the Clopper–Pearson 95% confidence interval are associated with $(x_0, x_1) = (1, 8)$ and $(x_0, x_1) = (2, 9)$, which are $0.4450 - 0.4439 = 0.5561 - 0.5550 = 0.0011$.

We define an ϵ-jump to correspond to a dwell time on an acceptance curve equal to 0. One example of an ϵ-jump is the point at $p = 0.5$ in Fig. 7. In that case, the actual coverage function stays on the acceptance curve $b(p, 1, 1)$ for a dwell time of 0. Some types of ϵ-jumps can cause problems. Allowing two consecutive upwards (downwards) ϵ-jumps results in confidence intervals for adjacent x that may have the same lower (upper) bounds, which we refer to as the "same bound" problem. Figure 10 shows the RMSE-minimizing 95% confidence intervals associated with $n = 10$. For these confidence intervals, the RMSE is 0.0162. As illustrated in Fig. 10, the confidence interval for $x = 5$ has the same lower bound as the confidence interval for $x = 6$. We would strongly prefer that these two confidence intervals not have the same lower bounds.

We now compare the RMSE of the RMSE-minimizing confidence interval with those of the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull intervals for small

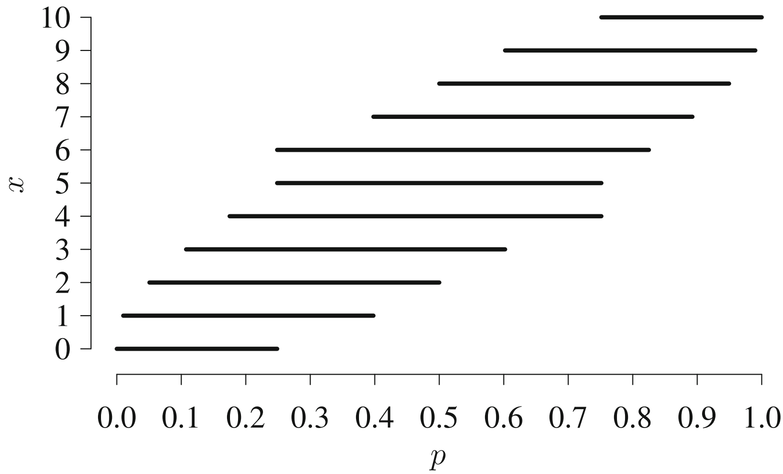

Fig. 10 RMSE-minimizing confidence intervals for $n = 10$

Table 5 RMSE comparison for $\alpha = 0.05$

n	Wilson–score	Jeffreys	Arcsine	Agresti–Coull	RMSE-minimizing
1	0.0641	0.0495	0.0499	0.0532	**0.0465**
2	0.0461	0.0392	0.0440	0.0460	**0.0387**
3	0.0366	0.0309	0.0380	0.0378	**0.0305**
4	0.0326	0.0376	0.0352	0.0316	**0.0261**
5	0.0295	0.0268	0.0334	0.0282	**0.0243**
6	0.0288	0.0309	0.0333	0.0256	**0.0222**
7	0.0241	0.0291	0.0300	0.0250	**0.0195**
8	0.0244	0.0287	0.0280	0.0239	**0.0212**
9	0.0237	0.0277	0.0272	0.0218	**0.0181**
10	0.0218	0.0243	0.0260	0.0216	**0.0162**
11	0.0220	0.0225	0.0263	0.0215	**0.0173**
12	0.0213	0.0235	0.0258	0.0203	**0.0157**

sample sizes and $\alpha = 0.05$. The RMSEs are displayed in the Table 5, which shows that our confidence interval achieves the lowest RMSE, which are set in boldface type, for $n = 1, 2, \ldots, 12$. Our confidence interval is the only one that has an RMSE below 0.02, for example, for $n = 9, 10, 11, 12$. The RMSE values are graphed in Fig. 11.

One solution to the same bound problem is to remove from consideration any symmetric Dyck paths that are associated with identical adjacent lower bounds. Figure 12 shows the 95% confidence intervals associated with $n = 6$ that satisfy this criterion. Although the same bound problem has been solved, the difference between the lower

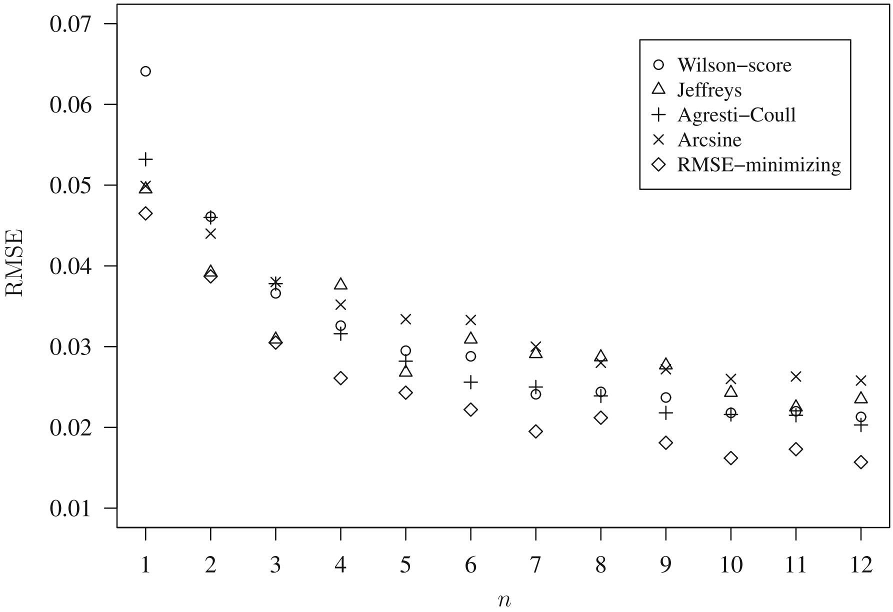

Fig. 11 RMSE comparison for $\alpha = 0.05$

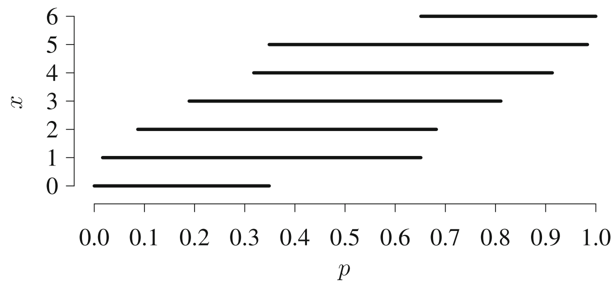

Fig. 12 RMSE-minimizing confidence intervals for $n = 6$ and $\alpha = 0.05$

bounds for $x = 4$ and $x = 5$ is quite small relative to the two adjacent differences. This prompted us to define a "smoothness" criterion for these confidence intervals, introduced in the next section, with a focus on the differences between the lower bounds.

6 Smoothed RMSE-minimizing confidence interval

In this section, we (a) define a measure of "smoothness" associated with a confidence interval for p, (b) design a smoothness constraint on the RMSE-minimizing confidence interval procedure, and (c) compare the smoothed RMSE-minimizing version of our confidence interval to other frequently-used non-conservative confidence intervals.

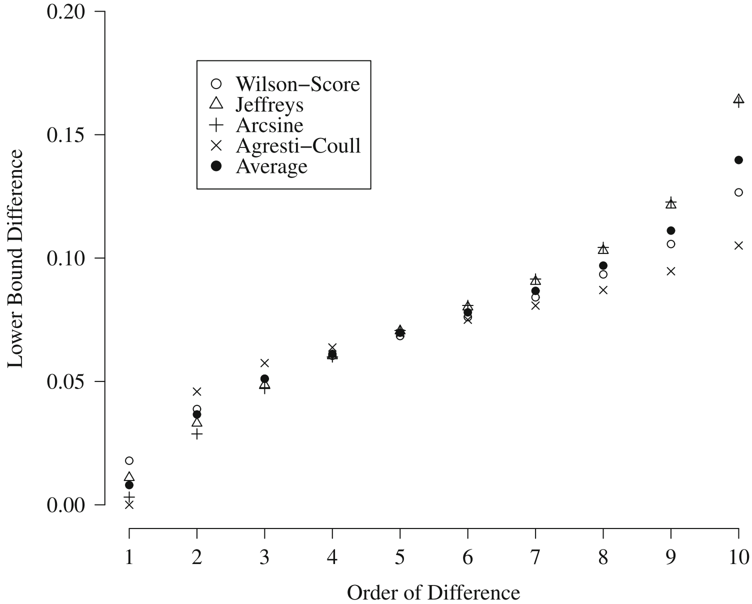

Fig. 13 Lower bound differences for $n = 10$ and $\alpha = 0.05$

By “smoothness,” we mean that the difference between two consecutive lower bounds is nondecreasing in x. For example, the difference between the lower bounds for $x = 2$ and $x = 3$ should be smaller than the difference between lower bounds for $x = 3$ and $x = 4$. Figure 13 contains a plot of the lower bound differences and their averages for the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull 95% confidence intervals for $n = 10$. There is a monotonically nondecreasing pattern of lower bound differences for all of the confidence interval procedures.

We propose a new metric for smoothness. Let $l_0, l_1, \ldots, l_n$ denote the lower confidence interval bounds for $x = 0, 1, \ldots, n$. Calculate the lower bound difference $d_k = l_k - l_{k-1}$ for $k = 1, 2, \ldots, n$. Calculate the ratio of two consecutive differences $r_k = d_{k+1}/d_k$ for $k = 1, 2, \ldots, n-1$. The *smoothness index* is defined by $\min\{r_1, r_2, \ldots, r_{n-1}\}$. If the smoothness index is greater than or equal to 1, which means the lower bound differences are non-decreasing, then the confidence interval procedure maintains the property of smoothness.

In order to avoid harmful ϵ-jumps and preserve the smoothness, we control the dwell time on each acceptance curve by placing lower and upper bounds on the dwell time. Since the RMSE-minimizing confidence interval is non-conservative, we employ four frequently-used non-conservative confidence intervals (Wilson–score, Jeffreys, Arcsine, and Agresti–Coull) to limit the dwell time associated with lower bound differences. Since the binomial confidence interval is symmetric, the dwell time bounds will automatically control upper bound differences as well. The dwell time constraints are defined by the following steps.

1. Compute the lower bounds for all four confidence interval procedures associated with sample size n. Denote each by l_{jk}, for $j = 1, 2, 3, 4$, which indexes the confi-

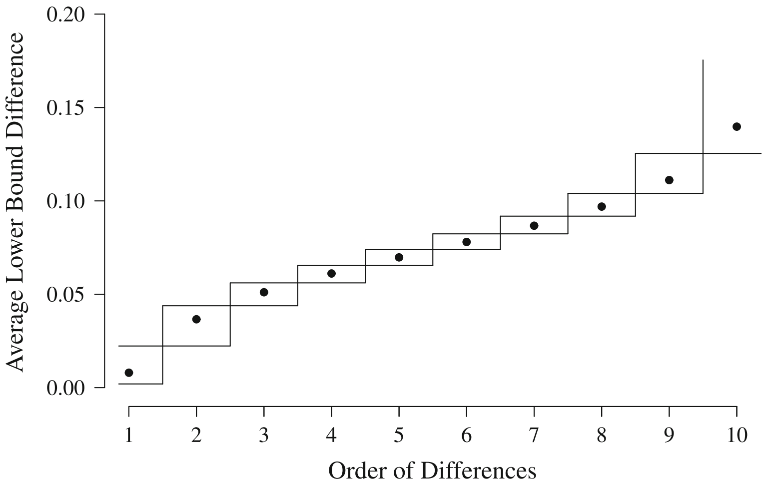

Fig. 14 Constraints on lower bound differences for $n = 10$ and $\alpha = 0.05$

dence interval procedure; $k = 0, 1, \ldots, n$, which indexes the number of successes x.

2. Calculate the averages of the four lower bounds $\bar{l_k} = \sum_{j=1}^{4} l_{jk}$, for $k = 0, 1, \ldots, n$.
3. Calculate the lower bound difference between two consecutive average lower bounds by $s_k = \bar{l_k} - \bar{l}_{k-1}$ for $k = 1, 2, \ldots, n$.
4. Set a lower limit for each lower confidence interval bound difference as

$$L_k = \begin{cases} (s_k - 0)/2 & k = 1 \\ (s_k - s_{k-1})/2 & k = 2, 3, \ldots, n. \end{cases}$$

5. Set an upper limit for each lower confidence interval bound difference as

$$U_k = \begin{cases} (s_{k+1} - s_k)/2 & k = 1, 2, \ldots, n-1 \\ 1 - \sum_{i=1}^{n-1} U_i & k = n. \end{cases}$$

6. Denote lower bounds in the smoothed RMSE-minimizing confidence interval for p associated with sample size n by $l_0, l_1, \ldots, l_n$. The subscripts correspond to x, the number of observed successes. Denote the difference between two consecutive lower bounds by $d_k = l_k - l_{k-1}$, for $k = 1, 2, \ldots, n$. Ensure that smoothness is achieved by requiring that $L_k \le d_k \le U_k$, for $k = 1, 2, \ldots, n$.

To illustrate the smoothing constraints for $n = 10$ and $\alpha = 0.05$, the black dots in Fig. 14 are the averages of the lower bound differences of the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull 95% confidence intervals. The solid horizontal lines are the constraints generated from following the steps above. We observe from Fig. 14 that the constraints are tighter in the middle and looser for x values at the extremes.

The algorithm with smoothing now works in two passes. It first calculates confidence interval bounds using the minimum-RMSE criterion. If the smoothness index is greater than or equal to one, the confidence interval is returned. If not, then the smooth

Table 6 Smoothed RMSE comparison for $\alpha = 0.05$

n	Wilson–score	Jeffreys	Arcsine	Agresti–Coull	RMSE-minimizing
1	0.0641	0.0495	0.0499	0.0532	**0.0465**
2	0.0461	0.0392	0.0440	0.0460	**0.0387**
3	0.0366	0.0309	0.0380	0.0378	**0.0305**
4	0.0326	0.0376	0.0352	0.0316	**0.0261**
5	0.0295	0.0268	0.0334	0.0282	**0.0243**
6	0.0288	0.0309	0.0333	0.0256	**0.0241**
7	0.0241	0.0291	0.0300	0.0250	**0.0211**
8	0.0244	0.0287	0.0280	0.0239	**0.0213**
9	0.0237	0.0277	0.0272	0.0218	**0.0207**
10	0.0218	0.0243	0.0260	0.0216	**0.0190**
11	0.0220	0.0225	0.0263	0.0215	**0.0198**
12	0.0213	0.0235	0.0258	0.0203	**0.0184**

symmetric Dyck path with the smallest RMSE is saved, and the problem is resolved with the additional smoothing constraints described above. Finally, the solution with the smaller RMSE is returned.

We compare the RMSE of our smoothed RMSE-minimizing confidence interval with those of the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull intervals for small sample sizes and $\alpha = 0.05$. The RMSE values are calculated in R and displayed in Table 6. The lowest RMSE value for each n is set in boldface type. Compared to the values in Table 5, some RMSE values increase because of the smoothness constraints. This reflects the trade-off between minimizing the RMSE value and achieving smoothness. However, Table 6 shows that our confidence interval still achieves the lowest RMSE for $n = 1, 2, \ldots, 12$, with no smoothing required for $n = 1, 2, \ldots, 5$.

Figure 15 contains three graphs that are associated with a sample size of $n = 10$ and a nominal coverage of $1-\alpha = 1-0.05 = 0.95$ for the RMSE-minimizing confidence interval procedure with constrained dwell time. The format is the same as that in Fig. 8. The top graph shows that although the RMSE-minimizing confidence interval does not have an $\epsilon-$jump, it is still possible that the dwell time on some acceptance curves to be very short. The middle graph shows that the confidence interval achieves smoothness because each lower bound difference is larger than the one before it. The bottom graph is consistent with the top graph in terms of the progression between acceptance curves associated with the optimal symmetric Dyck path.

7 Application

The binomial confidence interval is applied here to survival analysis. Consider the nonparametric estimation of the survivor function associated with the $n = 7$ rat survival times (in days) from Efron and Tibshirani (1993):

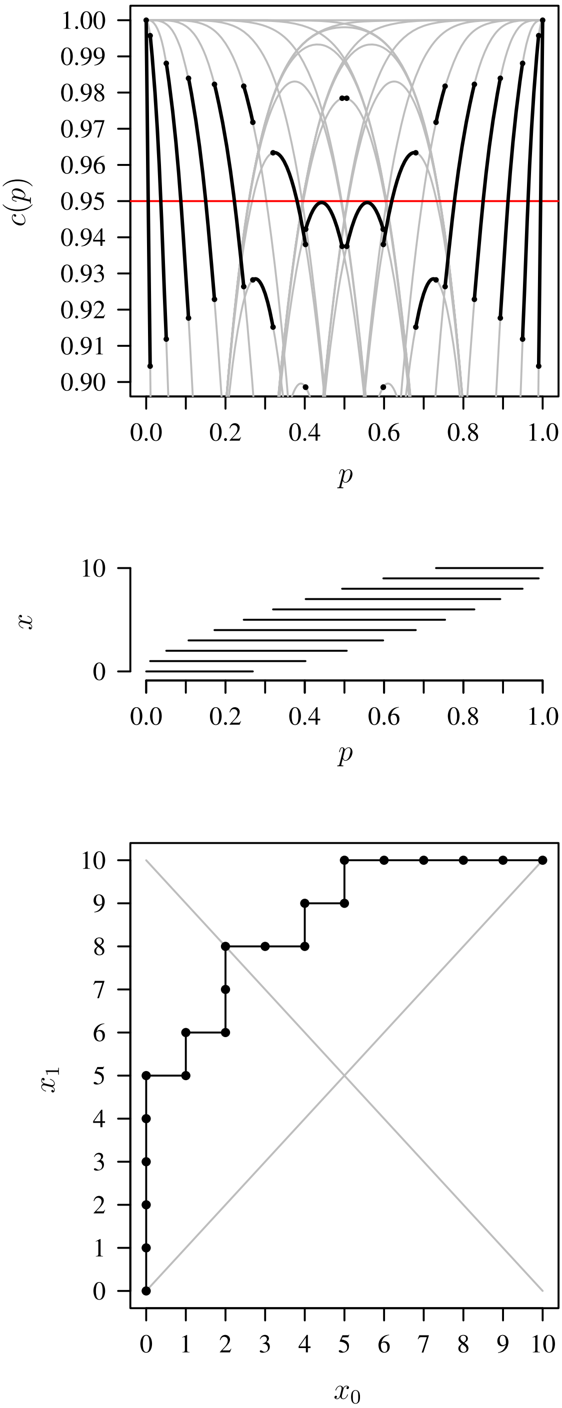

Fig. 15 Smooth RMSE-minimizing confidence intervals for $n = 10$ and $\alpha = 0.05$

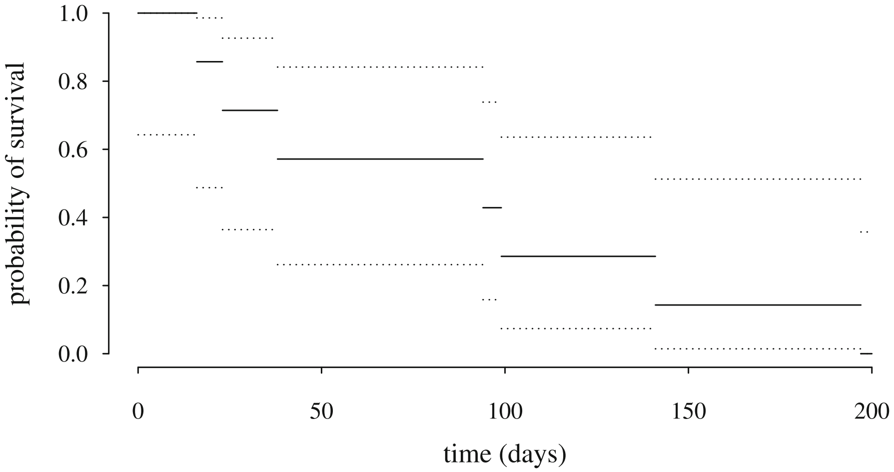

Fig. 16 Survivor function estimate for the rat survival data

16 23 38 94 99 141 197.

The empirical survival function, which takes a downward step of $1/n = 1/7$ at each data value, is given by the solid lines in Fig. 16. The dashed lines that denote 95% confidence intervals associated with the survival probability at any time are calculated using the smoothed RMSE-minimizing 95% confidence interval. These confidence bands are superior in their actual coverage to the usual confidence bands based on Greenwood's formula, which reduces to the Wald confidence interval for uncensored data. This has been confirmed by a Monte Carlo simulation experiment. The Wald confidence interval is notorious for poor actual coverage, as outlined by Brown et al. (2001).

Table 7 shows the results of six confidence interval procedures for calculating a 95% confidence interval for the probability of survival to 95 days. The Arcsine confidence interval is the widest of the intervals with a width of $0.787 - 0.118 = 0.669$ and the RMSE-minimizing confidence interval is the narrowest with a width of $0.739 - 0.159 = 0.58$. This 15% difference in confidence interval widths for the same data set can potentially result in differing conclusions. The $100(1-\alpha)\%$ confidence interval by Kim and Jang (2021) has bounds

$$p_a \pm z_{\alpha/2}\sqrt{\frac{p_c(1-p_c)}{n+d}},$$

where $p_a = (x+a)/(n+2a)$, $p_c = (x+c)/(n+2c)$, $a = 2.1 - 1/\sqrt{n}$, $c = 1.7 - 2/\sqrt{n}$, and $d = 4.0 - 1/\sqrt{n}$.

Table 7 Confidence intervals for $S(95)$ for $\alpha = 0.05$

Technique	Confidence interval
Wilson–score	$0.158 < S(95) < 0.750$
Jeffreys	$0.139 < S(95) < 0.765$
Arcsine	$0.118 < S(95) < 0.787$
Agresti–Coull	$0.158 < S(95) < 0.750$
Kim–Jang	$0.153 < S(95) < 0.751$
RMSE-minimizing	$0.159 < S(95) < 0.739$

8 Conclusions and further work

We have formulated and implemented an algorithm for calculating an approximate two-sided $100(1-\alpha)\%$ confidence interval for p from a random sample of n Bernoulli(p) observations that minimizes the root mean square error (RMSE) of the actual coverage function. We have added smoothness as a secondary constraint to avoid identical lower bounds for adjacent values of x. A smoothed RMSE-minimizing confidence interval has an actual coverage function that is closer to the nominal coverage than the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull confidence intervals for $n = 1, 2, \ldots, 12$. The marginal difference between the RMSE for the smoothed RMSE-minimizing confidence interval and the other confidence intervals decreases, although not necessarily monotonically in the sample size n.

This confidence interval procedure is implemented in the R function `binomTestMSE` in the `conf` package. For $n = 10, x = 3, \alpha = 0.05$, and smoothing, for example, the R statement

```
binomTestMSE(n = 10, x = 3, alpha = 0.05, smooth = 1)
```

returns the 95% confidence interval $0.107 < p < 0.598$. For other values of n and x, 95% confidence intervals $L < p < U$ are given in Table 8. For other α values, confidence intervals are given at `www.math.wm.edu/~leemis/tables`. The `binomTestMSE` function enumerates all symmetric Dyck paths for $n = 1, 2, \ldots, 15$ to achieve the smallest RMSE. The number of symmetric Dyck paths grows on the order of the factorial of n. The `binomTestMSE` function uses the symmetric Dyck paths associated with the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull confidence interval procedures with the smallest RMSE for $n \geq 16$ because of computation time constraints. Although there is no guarantee of optimality for $n \geq 16$, our experimentation showed that this approach typically results in the smallest RMSE. The structure of the algorithm is outlined in the Appendix.

We see two areas of future research. First, we would like to develop and implement an algorithm that allows the user to set a preferable smoothness on a continuous scale, using the smoothness index defined in Sect. 6. Second, we would like to develop the framework to minimize a weighted version of the RMSE which focuses on the most likely values of p based on expert opinion or previous data sets. This Bayesian framework would differ from the current framework which assigns equal weights to all values of p in (0, 1).

Table 8 Smoothed RMSE-minimizing 95% confidence intervals for p

n	x	L	U	n	x	L	U	n	x	L	U	n	x	L	U
1	0	0.000	0.900		3	0.136	0.691		11	0.741	1.000		10	0.454	0.881
	1	0.100	1.000		4	0.223	0.777	12	0	0.000	0.249		11	0.525	0.926
2	0	0.000	0.684		5	0.309	0.864		1	0.008	0.360		12	0.601	0.964
	1	0.050	0.950		6	0.419	0.936		2	0.042	0.453		13	0.684	0.993
	2	0.316	1.000		7	0.535	0.987		3	0.088	0.537		14	0.783	1.000
3	0	0.000	0.536		8	0.680	1.000		4	0.141	0.613	15	0	0.000	0.205
	1	0.033	0.811	9	0	0.000	0.306		5	0.193	0.683		1	0.007	0.299
	2	0.189	0.967		1	0.011	0.440		6	0.252	0.748		2	0.033	0.379
	3	0.464	1.000		2	0.057	0.551		7	0.317	0.807		3	0.069	0.452
4	0	0.000	0.500		3	0.120	0.647		8	0.387	0.859		4	0.110	0.519
	1	0.025	0.691		4	0.194	0.732		9	0.463	0.912		5	0.156	0.582
	2	0.135	0.865		5	0.268	0.806		10	0.547	0.958		6	0.201	0.642
	3	0.309	0.975		6	0.353	0.880		11	0.640	0.992		7	0.250	0.697
	4	0.500	1.000		7	0.449	0.943		12	0.751	1.000		8	0.303	0.750
5	0	0.000	0.401		8	0.560	0.989	13	0	0.000	0.235		9	0.358	0.799
	1	0.020	0.599		9	0.694	1.000		1	0.008	0.339		10	0.418	0.844
	2	0.106	0.766	10	0	0.000	0.269		2	0.039	0.427		11	0.481	0.890
	3	0.234	0.894		1	0.010	0.402		3	0.081	0.507		12	0.548	0.931
	4	0.401	0.980		2	0.051	0.506		4	0.129	0.580		13	0.621	0.967
	5	0.599	1.000		3	0.107	0.598		5	0.177	0.648		14	0.701	0.993

Table 8 continued

n	x	L	U	n	x	L	U	n	x	L	U	n	x	L	U
6	0	0.000	0.412		4	0.172	0.680		6	0.231	0.711		15	0.795	1.000
	1	0.017	0.579		5	0.246	0.754		7	0.289	0.769	16	0	0.000	0.195
	2	0.087	0.708		6	0.320	0.828		8	0.352	0.823		1	0.006	0.284
	3	0.189	0.811		7	0.402	0.893		9	0.420	0.871		2	0.031	0.361
	4	0.292	0.913		8	0.494	0.949		10	0.493	0.919		3	0.065	0.430
	5	0.421	0.983		9	0.598	0.990		11	0.573	0.961		4	0.103	0.495
	6	0.588	1.000		10	0.731	1.000		12	0.661	0.992		5	0.145	0.556
7	0	0.000	0.357	11	0	0.000	0.259		13	0.765	1.000		6	0.187	0.613
	1	0.014	0.513		1	0.009	0.377	14	0	0.000	0.217		7	0.233	0.667
	2	0.074	0.636		2	0.046	0.475		1	0.007	0.316		8	0.281	0.719
	3	0.159	0.739		3	0.096	0.562		2	0.036	0.399		9	0.333	0.767
	4	0.261	0.841		4	0.155	0.642		3	0.074	0.475		10	0.387	0.813
	5	0.364	0.926		5	0.220	0.714		4	0.119	0.546		11	0.444	0.855
	6	0.487	0.986		6	0.286	0.780		5	0.168	0.611		12	0.505	0.897
	7	0.643	1.000		7	0.358	0.845		6	0.217	0.672		13	0.570	0.935
8	0	0.000	0.320		8	0.438	0.904		7	0.270	0.730		14	0.639	0.969
	1	0.013	0.465		9	0.525	0.954		8	0.328	0.783		15	0.716	0.994
	2	0.064	0.581		10	0.623	0.991		9	0.389	0.832		16	0.805	1.000

Acknowledgements The authors thank Hayeon Park, Chris Weld, two anonymous reviewers, and the editor for their helpful comments concerning this paper.

Appendix

The structure of the algorithm for constructing the $100(1-\alpha)\%$ RMSE-minimizing confidence interval is given below. The four existing confidence interval procedures used in the smoothing and the large-sample portions of the algorithm are the Wilson–score, Jeffreys, Arcsine, and Agresti–Coull.

```
input  : n: sample size
         x: number of successes
         α: confidence interval significance level
         smooth: 0 for no smoothing; 1 for smoothing
output: lower and upper 100(1 − α)% confidence interval bounds (CIBs) for p
if n ≤ 15 then
    generate all npt = (n choose ⌊n/2⌋) symmetric Dyck paths
    for i from 1 to npt do
        for j from 1 to n do
            numerically solve for CIBs that minimize the RMSE
    if smooth = 1 then
        perform smoothing
else
    determine symmetric Dyck paths for the four existing confidence intervals
    for i from 1 to 4 do
        for j from 1 to n do
            numerically solve for CIBs that minimize the RMSE
    if smooth = 1 then
        perform smoothing
return the CIBs with the smallest RMSE
```

This algorithm has been implemented in R in the `binomTestMSE` function in the `conf` package, which consists of about 700 lines of code. The numerical solution that minimizes the RMSE is performed by the `uniroot.all` function in R.

References

Agresti A, Coull BA (1998) Approximate is better than ‘exact’ for interval estimation of binomial proportions. Am Stat 52:119–126

Anscombe FJ (1956) On estimating binomial response relations. Biometrika 43:461–464

Balch MS (2020) New two-sided confidence intervals for binomial inference derived using Walley’s imprecise posterior likelihood as a test statistic. Int J Approx Reason 123:77–98

Blaker H (2000) Confidence curves and improved exact confidence intervals for discrete distributions. Can J Stat 28:783–798

Brown LD, Cai TT, DasGupta A (2001) Interval estimation for a binomial proportion. Stat Sci 16:101–117

Clopper CJ, Pearson ES (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404–413

Deng LH, Deng EY, Shapiro LW (2009) The Riordan group and symmetric lattice paths. arXiv preprint. arXiv:0906.1844

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. CRC Press, Boca Raton

Kása Z (2009) Generating and ranking of Dyck words. Acta Universitatis Sapientiae Informatica 1:109–118

Kim J, Jang W (2021) A generalized Agresti–Coull type confidence interval for a binomial proportion. J Korean Stat Soc. https://doi.org/10.1007/s42952-021-00143-3

Lyles RH, Weiss P, Waller LA (2020) Calibrated Bayesian credible intervals for binomial proportions. J Stat Comput Simul 90:75–89

Meeker WQ, Hahn GJ, Escobar LA (2017) Statistical intervals: a guide for practitioners and researchers, 2nd edn. Wiley, Hoboken

Park H, Leemis LM (2019) Ensemble confidence intervals for binomial proportions. Stat Med 38:3460–3475

Wilcox R (2020) A note on inferences about the probability of success. J Mod Appl Stat Methods 18(1), Article 30. https://doi.org/10.22237/jmasm/1556670420

Wilson EB (1927) Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22:209–212

Yaacoub T, Goldsman D, Mei Y, Moustakides GV (2019) Tandem-width sequential confidence intervals for a Bernoulli proportion. Seq Anal 38:163–183